Basic Setup and Configuration

Guidelines and Restrictions for Deploying APIC Cluster Connectivity to the Fabric Over a Layer 3 Network

When deploying a layer 3 connected APIC cluster, follow these guidelines and limitations.

- You can use the infra VLAN to connect to the IPN router or you can use a different VLAN.
- APIC Cluster Connectivity to the Fabric Over a Layer 3 Network can be configured only for a new APIC cluster. An existing APIC cluster can be converted to a layer 3 connected APIC cluster only after erasing its setup.
- All APIC cluster sizes are supported in a layer 3 connected APIC pod.
- APICs in a layer 3 connected APIC pod cannot form a cluster with APICs within the fabric pod. In this topology, there should be no APICs in the fabric pod.
- The layer 3 connected APICs can be in the same subnet or in different subnets.
- The layer 3 connected APICs can be geographically distributed from each other provided that the latency between APICs and with the fabric pod does not exceed 50 milliseconds round-trip time (RTT), which translates approximately to a geographical distance of up to 2,500 miles.
- Although any device that can meet the IPN network requirements can be used as an IPN device, we recommend to deploy, when possible, switches of the Cisco Nexus 9300 Cloud Scale family. These are the devices most commonly found in production and also the devices more frequently validated in Cisco internal testing. For further information about IPN device requirements, see "Inter-Pod Connectivity Deployment Considerations" in the ACI Multi-Pod White Paper.

- The APIC subnets must be advertised to the spines as either OSPF or BGP routes. An OSPF underlay is supported for all supported releases. A BGP underlay is supported with APIC release 5.2(3) and later releases.
- Because all control plane traffic between the APIC cluster and the fabric pod traverses the IPN, we recommend configuring QoS for this traffic. For specific recommendations, see Configuring QoS for the Layer 3 Connected APIC Cluster, on page 12.
- APIC Cluster Connectivity to the Fabric Over a Layer 3 Network does not support the following:
  - ACI Virtual Edge (AVE)
  - ACI CNI for Kubernetes (Redhat Openshift, SUSE/Rancher RKE, Upstream Kubernetes on Ubuntu)
  - ACI ML2 for Openstack (Redhat Openstack, Canonical Openstack)
- APIC Cluster Connectivity to the Fabric Over a Layer 3 Network does not support vAPIC or vPOD.
- APIC Cluster Connectivity to the Fabric Over a Layer 3 Network supports standby APIC.
- APIC Cluster Connectivity to the Fabric Over a Layer 3 Network supports strict mode. In strict mode, you must approve the controller explicitly.

Forming the APIC Cluster

Provisioning the APIC Cluster-Facing IPN Device

This section describes the configuration of the IPN device connected to Pod 0, the APIC cluster pod. In the figure in About APIC Cluster Connectivity to the Fabric Over a Layer 3 Network, the cluster-facing IPN device is shown as IPN0. As a recommended practice, the IPN0 example comprises two devices for redundancy. The fabric interface of each APIC is dual-homed to the two devices. In the following configuration example, two Cisco Nexus 9300 series switches (IPN0a and IPN0b) are configured with the following choices:

- VLAN 1500 is configured for the APIC interface.
- The switch interfaces are configured as layer 2 trunk ports. As an alternative, the interfaces could be access ports if the APIC fabric interface is configured to use VLAN 0 during APIC setup.
- Both switches are configured using HSRP to share a single IP address that serves as the APIC subnet default gateway address.
- APIC subnets are advertised to the spines using OSPF as the underlay protocol. As an alternative, a BGP underlay could be deployed if the APIC release is 5.2(3) or later.

```
# Example configuration of IPN0a:

interface Vlan1500
  no shutdown
  vrf member IPN
  ip address 172.16.0.252/24
  ip ospf passive-interface
  ip router ospf 1 area 0.0.0.0
  hsrp version 2
  hsrp 1500
```

```
    ip 172.16.0.1

interface Ethernet1/1
  switchport mode trunk
  switchport trunk native vlan 1500
  spanning-tree port type edge trunk


# Example configuration of IPN0b:

interface Vlan1500
  no shutdown
  vrf member IPN
  ip address 172.16.0.253/24
  ip ospf passive-interface
  ip router ospf 1 area 0.0.0.0
  hsrp version 2
  hsrp 1500
    ip 172.16.0.1

interface Ethernet1/1
  switchport mode trunk
  switchport trunk native vlan 1500
  spanning-tree port type edge trunk
```

Initial APIC Cluster Setup

When an APIC is launched for the first time, the APIC console presents a questionnaire with a series of initial setup options. Beginning with Cisco APIC Release 5.2(1), the questionnaire asks whether you are installing a layer 3 connected APIC cluster cluster (a "Standalone APIC Cluster"), followed by a set of questions specific to that scenario.

For each APIC, you must provide the APIC's interface VLAN, the APIC's IP address, and the APIC's gateway IP address. The APICs can be in the same subnet or in different subnets. For a layer 3 connected APIC cluster, the questionnaire does not ask for the Pod ID (it will be Pod 0 by default) or for the TEP pool.

In the initial configuration of the first APIC in the layer 3 connected APIC cluster, you must set the controller ID to 1, which causes the questionnaire to ask whether this is the first APIC in the cluster.

For the first APIC, the questionnaire asks the following questions.

Note In the following examples, the questions related to deploying APIC cluster connectivity to the fabric over a layer 3 network are indicated with asterisks. (These asterisks do not appear in the actual command lines.)

```
Cluster configuration...
  Enter the fabric name [ACI Fabric1]:
  Enter the fabric ID (1-128) [1]:
  Enter the number of active controllers in the fabric (1-9) [3]:
  Is this a standby controller? [NO]:
  Enter the controller ID (1-3) [1] 1
* Standalone APIC Cluster ? yes/no [no]: yes
* Enter the VLAN ID for interface (0-4094): 0
* Enter the APIC IPV4 address [A.B.C.D/NN]: 172.16.0.2/24
* Enter the IPV4 address of the APIC default gateway [A.B.C.D]: 172.16.0.1
* First APIC in the Cluster ? yes/no [yes]: yes
```

```
  Enter the controller name [apic1]:
...
```

As you bring up additional APICs, with controller IDs other than 1, the questionnaire asks for the IP address of an active APIC already in the cluster. For APIC 2 and APIC 3, you must provide the IP address of APIC 1, as shown in the following example. For any APICs beyond APIC 3, you can provide the IP address of any active APIC already in the cluster.

```
Cluster configuration...
  Enter the fabric name [ACI Fabric1]:
  Enter the fabric ID (1-128) [1]:
  Enter the number of active controllers in the fabric (1-9) [3]:
  Is this a standby controller? [NO]:
  Enter the controller ID (1-3) [1]: 2
* Standalone APIC Cluster ? yes/no [no]: yes
* Enter the VLAN ID for interface (0-4094): 1500
* Enter the APIC IPV4 address [A.B.C.D/NN]: 172.16.0.3/24
* Enter the IPV4 address of the APIC default gateway [A.B.C.D]: 172.16.0.1
* Enter the IPV4 address of an active APIC [A.B.C.D]: 172.16.0.2
  Enter the controller name [apic2]:
...
```

When the APIC is up, you can check the LLDP neighbors using the following APIC CLI command while logged in as `admin`:

```
[admin@apic2 ~]# show lldp
```

For general details about the initial setup questionnaire, see the "Setup for Active and Standby APIC" section in the *Cisco APIC Getting Started Guide*.

For additional, less commonly needed cluster procedures, see Cluster Management Tasks, on page 11.

After creating the layer 3 connected APIC cluster, you can configure the APIC to communicate over the inter-pod network (IPN) with the fabric pod, as described in Preparing Connectivity to the Fabric Pod, on page 6.

Discovering the Fabric

Provisioning the Fabric-Facing IPN Device

This section describes the configuration of the MPod IPN, which is the IPN device connected to a fabric pod. The IPN is not managed by the APIC. It must be preconfigured with the following information:

- Configure the interfaces connected to the spines of the fabric pod (Pod 1). Use Layer 3 sub-interfaces tagging traffic with VLAN-4 and increase the MTU at least 50 bytes above the maximum MTU required for inter-site control plane and data plane traffic.
- Enable OSPF on the sub-interface specifying the OSPF process and area ID.

Note With APIC release 5.2(3) or later, you have the option of enabling BGP instead of OSPF.

- Enable DHCP Relay on the IPN interfaces connected to spines.
- Enable PIM.
- Add bridge domain GIPo range as PIM Bidirectional (**bidir**) group range (default is 225.0.0.0/15).

  A group in **bidir** mode has only shared tree forwarding capabilities.
- Add 239.255.255.240/28 as PIM **bidir** group range.
- Enable PIM on the interfaces connected to all spines.

Note Multicast is not required for a single pod fabric with a layer 3 connected APIC cluster, but it is required between pods in a multi-pod fabric.

Note When deploying PIM **bidir**, at any given time it is only possible to have a single active RP (Rendezvous Point) for a given multicast group range. RP redundancy is hence achieved by leveraging a **Phantom RP** configuration. Because multicast source information is no longer available in Bidir, the Anycast or MSDP mechanism used to provide redundancy in sparse-mode is not an option for **bidir**.

Example: Fabric-Facing IPN Switch Configuration

The following switch configuration example is for a Cisco Nexus 9300 series switch deployed as the MPod IPN, which is IPN 1 in the figure in About APIC Cluster Connectivity to the Fabric Over a Layer 3 Network. The DHCP relay configuration allows the fabric to be discovered by the APIC cluster.

Note The deployment of a dedicated VRF in the IPN for inter-pod connectivity is optional, but is a best practice recommendation. As an alternative, you can use a global routing domain.

```
feature dhcp
feature pim
service dhcp
ip dhcp relay

# Create a new VRF.
vrf context overlay-1
  ip pim rp-address 12.1.1.1 group-list 225.0.0.0/15 bidir
  ip pim rp-address 12.1.1.1 group-list 239.255.255.240/28 bidir

interface Ethernet1/54.4
  mtu 9150
  encapsulation dot1q 4
  vrf member overlay-1
  ip address 192.168.0.1/30
  ip ospf network point-to-point
  ip router ospf infra area 0.0.0.0
  ip dhcp relay address 172.16.0.2
  ip dhcp relay address 172.16.0.3
  ip dhcp relay address 172.16.0.4
  no shutdown
```

```
interface loopback29
  vrf member overlay-1
  ip address 12.1.1.2/30

router ospf infra
  vrf overlay-1
    router-id 29.29.29.29
```

In the preceding example, the underlay protocol is OSPF. The following example shows a BGP underlay configuration:

```
router bgp 65010
  vrf IPN
    neighbor 192.168.0.2 remote-as 65001
      address-family ipv4 unicast
        disable-peer-as-check
```

In the BGP configuration, the `disable-peer-as-check` command is needed for multi-pod because each pod uses the same ASN.

Preparing Connectivity to the Fabric Pod

Before bringing up the fabric pod (Pod 1), you first must pre-configure the layer 3 connected APIC cluster (Pod 0) for connectivity through the IPN to a spine in the fabric pod. This is necessary for automatic fabric discovery.

Before you begin

- If the layer 3 connected APIC cluster is deployed in a separate security zone from the fabric, configure the firewall to allow any necessary protocols and ports shown in APIC Protocols and Ports, on page 9.
- Configure the inter-pod network (IPN) device that is connected to the fabric pod spines.
- Configure a fabric external routing profile.
- Configure an OSPF interface policy if you are using OSPF as the underlay protocol.

Step 1 Log in to one of the APICs in the layer 3 connected cluster.

Step 2 Choose **Fabric** > **Inventory** > **Pod Fabric Setup Policy**.

Step 3 In the work pane, click the **Physical Pods** tab and click the **+** symbol.

The **Set Up Pod TEP Pool** dialog box opens.

Step 4 In the **Set Up Pod TEP Pool** dialog box, complete the following steps:

a) Using the **Pod ID** selector, choose Pod 1.
b) In the **TEP Pool** field, enter the TEP pool of the fabric pod.
c) Click **Submit**.

Step 5 In the navigation pane, expand **Quick Start** and click **Add Pod**.

Step 6 In the work pane, click **Add Pod**.

Step 7 In the **Configure Interpod Connectivity STEP 1 > Overview** panel, review the tasks that are required to configure interpod network (IPN) connectivity, and then click **Get Started**.

Step 8 In the **Configure Interpod Connectivity STEP 2 > IP Connectivity** dialog box, complete the following steps:

a) If you see a **Name** field in an **L3 Outside Configuration** area, choose an existing fabric external routing profile from the **Name** drop-down list.

b) Using the **Spine ID** selector, choose one spine in Pod 1 that will be the initial spine to communicate with APIC 1 in Pod 0.

c) In the **Interfaces** area, in the **Interface** field, enter the spine switch interface (slot and port) used to connect to the IPN.

Click the **+** (plus) icon to add more interfaces.

d) In the **IPV4 Address** field, enter the IPv4 gateway address and network mask for the interface.

e) From the **MTU (bytes)** drop-down list, choose a value for the maximum transmit unit of the external network.

The MTU should be 9150 (the default). This value should also be configured on the IPN interface.

f) Click **Next**.

Step 9 In the **Configure Interpod Connectivity STEP 3 > Routing Protocols** dialog box, in the **OSPF** area, complete the following steps to configure OSPF for the spine to IPN interface:

a) Leave the **Use Defaults** checked or uncheck it.

When the **Use Defaults** check box is checked, the GUI conceals the optional fields for configuring Open Shortest Path (OSPF). When it is unchecked, it displays all the fields. The check box is checked by default.

b) In the **Area ID** field, enter the OSPF area ID.

c) In the **Area Type** area, choose an OSPF area type.

You can choose **NSSA area** or **Regular area** (the default). **Stub area** is not supported.

d) (Optional) With the **Area Cost** selector, choose an appropriate OSPF area cost value.

e) From the **Interface Policy** drop-down list, choose or configure an OSPF interface policy.

You can choose an existing policy, or you can create one with the **Create OSPF Interface Policy** dialog box. An example is shown in the following table:

Table 1: OSPF Interface Policy Example

Property	Setting
Name	ospfIfPol
Network Type	Point-to-point
Priority	1
Cost of Interface	unspecified
Interface Controls	none checked
Hello Interval (sec)	10
Dead Interval (sec)	40
Retransmit Interval (sec)	5

Step 10 In the **Configure Interpod Connectivity STEP 3 > Routing Protocols** dialog box, in the **BGP** area, leave the **Use Defaults** checked or uncheck it.

The **Use Defaults** check box is checked by default. When the check box is checked, the GUI conceals the fields for configuring Border Gateway Protocol (BGP). When it is unchecked, it displays all the fields. If you uncheck the box, configure the following steps:

a) Leave the **Use Defaults** checked or uncheck it.

b) In the **Community** field, enter the community name.

We recommend that you use the default community name. If you use a different name, follow the same format as the default.

c) In the **Peering Type** field, choose either **Full Mesh** or **Route Reflector** for the route peering type.

If you choose **Route Reflector** in the **Peering Type** field and you later want to remove the spine switch from the controller, you must first disable **Route Reflector** in the *BGP Route Reflector* page. Not doing so results in an error.

To disable a route reflector, right-click on the appropriate route reflector in the **Route Reflector Nodes** area in the **BGP Route Reflector** page and select **Delete**. See the section "Configuring an MP-BGP Route Reflector Using the GUI" in the chapter "MP-BGP Route Reflectors" in the *Cisco APIC Layer 3 Networking Configuration Guide*.

d) In the **Peer Password**, field, enter the BGP peer password. In the **Confirm Password** field, reenter the BGP peer password.

e) In the **Route Reflector Nodes** area, click the **+** (plus) icon to add nodes.

For redundancy purposes, more than one spine is configured as a route reflector node: one primary reflector and one secondary reflector. It is best practice to deploy at least one external route reflector per pod for redundancy purposes.

The **External Route Reflector Nodes** fields appear only if you chose **Route Reflector** as the peering type.

Step 11 Click **Next**.

Step 12 In the **Configure Interpod Connectivity STEP 4 > External TEP** dialog box, complete the following steps:

a) Leave the **Use Defaults** checked or uncheck it.

The **Use Defaults** check box is checked by default. When the check box is checked, the GUI conceals the optional fields for configuring the external TEP pool. When it is unchecked, it displays all the fields.

b) Note the nonconfigurable values in the **Pod** and **Internal TEP Pool** fields.

c) In the **External TEP Pool** field, enter the external TEP pool for the physical pod.

The external TEP pool must not overlap the internal TEP pool or external TEP pools belonging to other pods.

d) In the **Data Plane TEP IP** field, accept the default, which is generated when you configure the **External TEP Pool**; if you enter another address, it must be outside of the external TEP pool.

Step 13 Click **Next**.
The **Summary** panel appears, displaying a list of policies created by this wizard. You can change the names of these policies here.

Step 14 Click **Finish**.

What to do next

Monitor the discovery and registration of the fabric nodes by APIC, as summarized in the following section.

Summary of Fabric Discovery and Registration

The following is a summary of the switch registration and discovery process.

- The IPN, acting as a DHCP relay agent, forwards DHCP requests from the spine to the APIC.
- The APIC allocates an IP address for the spine interface to the IPN and a TEP IP from the TEP pool configured for the fabric pod containing the spine. At this point, the spine joins the ACI fabric.
- The spine advertises the TEP subnet to the IPN through OSPF or BGP so that the IPN learns this subnet.
- The spine acts as a DHCP relay agent for its connected leaf switches, and forwards the requests to the APIC.
- The leaf switches are then visible to the APIC and appear in the section of **Nodes Pending Registration** in **Fabric > Inventory > Fabric Membership**.
- You must manually register these leaf switches on the APIC.
- After a leaf switch is registered, the APIC forwards the TEP IP address and DHCP configuration information to the leaf, which then joins the ACI fabric. Through this discovery process, the layer 3 connected APIC cluster discovers all switches in the fabric pod.

APIC Protocols and Ports

The following table lists the protocols and port numbers used for communications between APIC and the fabric switches. When deploying a layer 3 connected APIC cluster behind a firewall, be sure to allow the protocols needed by your ACI fabric.

Service	Source	Destination	Protocol	Destination Port (range)
Kafka Message Broker	Switch	APIC	tcp	9092
Kafka server	Switch	APIC	tcp	9094
Data Ingestion Receiver port for external access	Switches, APIC	SE Kafka Broker	tcp	30001
NI Tech Support SCP	Switches, APIC	SE App SSHD	tcp	2022
APIC Internal SSHD	Switches	APIC	tcp	1022
Streaming Telemetry	Switches	SE	udp	5640 - 5656
http protocol over TLS/SSL	ND, APIC, ND Insights, Switches	ND, APIC, ND Insights, Switches		443
SSH	ND Insights/APIC	Switches	tcp	22

Service	Source	Destination	Protocol	Destination Port (range)
IFM	Switches, APIC	Switches, APIC	tcp	12006 – 13541
NGINX	Switch, APIC, and SE Infra	APIC	tcp	7777
DHCP	Switches	APIC	udp	67
DHCP	Switches	APIC	udp	67
DHCP	Switches	APIC	udp	68
NTP	Switches	APIC	udp	123
http protocol over TLS/SSL	ND, APIC, ND Insights, Switches	ND, APIC, ND Insights, Switches		443
SSH	ND Insights	Switches	tcp	22
OSPF	Switches	Switches	tcp	89
ISIS	Switches	Switches	ip	ip protocol 124
ISIS	Switches	Switches	udp, tcp	2042
ISIS	Switches	Switches	udp, tcp	2043
VXLAN/iVXLAN	Switches	Switches	udp	4789
Python	APIC	APIC	udp, tcp	4646
Nomad	APIC	APIC	udp, tcp	4646, 4647,4649
Java	APIC	APIC	udp, tcp	9072
eforward	APIC	APIC	udp, tcp	2181
spcsdlobby	APIC	APIC	tcp	2888
glusterd	APIC	APIC	tcp	24007
Consul	APIC	APIC	udp, tcp	8500
fmtp (Flight Message Transfer Protocol)	APIC	APIC	udp, tcp	8500
vrace (Virtual Racing Service)	APIC	APIC	udp, tcp	9300

Cluster Management Tasks

This section contains APIC cluster management tasks that are not needed in normal operation but may be necessary in some situations.

Replacing APIC 1

In the special case of bringing up a replacement for APIC 1, you can indicate that, although the controller ID is 1, it is not the first APIC in a new cluster. The APIC will then ask for the IP address of another active APIC already in the cluster, with which it will sync up and join the cluster. The questionnaire flow for this case is shown in the following example.

```
Cluster configuration...
  Enter the fabric name [ACI Fabric1]:
  Enter the fabric ID (1-128) [1]:
  Enter the number of active controllers in the fabric (1-9) [3]:
  Is this a standby controller? [NO]:
  Enter the controller ID (1-3) [1] 1
* Standalone APIC Cluster ? yes/no [no]: yes
* Enter the APIC IPV4 address [A.B.C.D/NN]: 172.16.0.2/24
* Enter the IPV4 address of the APIC default gateway [A.B.C.D]: 172.16.0.1
* First APIC in the Cluster ? yes/no [yes]: no
* Enter the IPV4 address of an active APIC [A.B.C.D]: 172.16.0.3
  Enter the controller name [apic1]:
...
```

For more information about the initial setup questionnaire, see Initial APIC Cluster Setup, on page 3.

Clean Rebooting the APICs

If it becomes necessary to clean reboot one or more APICs in the cluster, run the following APIC CLI commands in admin mode.

To clean reboot APIC 1, run the following commands:

```
[admin@apic1 ~]# acidiag touch clean
[admin@apic1 ~]# acidiag touch cleanactiveapiclist
[admin@apic1 ~]# acidiag reboot
```

To clean reboot other APICs, run the following commands:

```
[admin@apic2 ~]# acidiag touch clean
[admin@apic2 ~]# acidiag reboot
```

For more information about the `acidiag` CLI command, see the *Cisco APIC Troubleshooting Guide*.

Upgrading Standby APICs

Standby APICs are supported in the layer 3 connected APIC architecture. The normal policy upgrade will upgrade only the active APICs in the layer 3 connected APIC cluster. Follow the steps in this section to upgrade the standby APICs and to add the standby APIC subnet in the fabric external routing profile of the infra tenant:

1. On the menu bar, choose **Tenants** > **infra**.

2. In the Navigation pane, expand **infra** > **Policies** > **Protocol**.
3. Right-click **Fabric Ext Connection Policies** and choose **Create Intrasite/Intersite Profile**.
4. Click the + symbol on **Fabric External Routing Profile**.
5. Under **Fabric External Routing Profile**, add the subnet of the standby APIC.
6. Click **Update**, then click **Submit**.

Replacing an Active APIC with a Standby APIC

When you replace an active APIC (APIC 3, for example) with a standby APIC (APIC 23, for example), the formerly standby APIC (APIC 23) becomes one of the new active APICs and the formerly active APIC (APIC 3) is shut down.

Follow these steps to add the subnet of the formerly standby APIC (APIC 23) into the **Fabric Ext Connection Policies**:

1. On the menu bar, choose **Tenants** > **infra**.
2. In the Navigation pane, expand **infra** > **Policies** > **Protocol** > **Fabric Ext Connection Policies**.
3. If no Fabric Ext Connection Policy exists under **Fabric Ext Connection Policies**, skip to the next step. If a Fabric Ext Connection Policy already exists, click the policy to open it in the work pane. Perform the following substeps and then skip Step 4:
   a. Click the + symbol on **Fabric External Routing Profile**.
   b. Under **Fabric External Routing Profile**, add the subnet of the formerly standby APIC (APIC 23).
   c. Click **Update**, then click **Submit**.
4. If no Fabric Ext Connection Policy exists under **Fabric Ext Connection Policies**, create a policy using the following substeps:
   a. Click **Fabric Ext Connection Policies** and choose **Create Intrasite/Intersite Profile**.
   b. Click the + symbol on **Fabric External Routing Profile**.
   c. Under **Fabric External Routing Profile**, add the subnet of the formerly standby APIC (APIC 23).
   d. Click **Update**, then click **Submit**.

Configuring QoS for the Layer 3 Connected APIC Cluster

Because all traffic between the APIC cluster and the fabric traverses the IPN, we recommend configuring QoS for this traffic. Specific recommendations are as follows:

Note The configuration examples in this section are for a Cisco Nexus 9300 series switch used as an IPN device. The configuration may differ when using a different platform.

- Enable a DSCP class CoS translation policy.

- Retain end-to-end DCSP values for all IPN traffic. When the DSCP class CoS translation policy is enabled, the spine switches will set the DSCP value in packets sent to the IPN per this policy. Traffic destined to the APICs will use the Policy Plane class. Inter-pod control plane traffic (that is, OSPF and MP-BGP packets) will use the Control Plane class. Ensure that the Policy Plane class is configured for Expedited Forwarding (EF) and the Control Plane class is configured for CS4. The IPN network can be configured to prioritize these two classes to ensure that the policy plane and control plane remain stable during times of congestion.

  The following example shows the configuration for setting QoS for policy plane traffic on a Cisco Nexus 9300 series switch used as an IPN device.

  ```
  interface Vlan1500
    no shutdown
    vrf member overlay-1
    ip address 172.16.0.2/24
    ip ospf passive-interface
    ip router ospf 1 area 0.0.0.0
    hsrp version 2
    hsrp 100
      ip 172.16.0.1

  ip access-list APIC_Cluster
    10 permit ip 172.16.0.0/24 any

  class-map type qos match-all APIC-Class
    match access-group name APIC_Cluster
  policy-map type qos APIC_Policy
    class APIC-Class
      set dscp 46

  interface Ethernet1/1
    switchport mode trunk
    switchport access vlan 1500
  service-policy type qos input APIC_Policy

  interface Ethernet1/2
    switchport mode trunk
    switchport access vlan 1500
  service-policy type qos input APIC_Policy

  interface Ethernet1/3
    switchport mode trunk
    switchport access vlan 1500
  service-policy type qos input APIC_Policy
  ```

- We also recommend that you rate limit non-policy-plane traffic to the APIC to prevent drops on policy plane traffic. Deploy a policer on the cluster-facing IPN to limit traffic with a DSCP value other than 46 (EF) to 4 Gbps with burst of 60 Mbps. The following example shows a rate limiting configuration on a Cisco Nexus 9300 series switch used as an IPN device.

  ```
  class-map type qos match-all no_rate_limit
    match dscp 46

  policy-map type qos APIC_Rate_Limit
    class no_rate_limit
    class class-default
      police cir 4 gbps bc 7500 kbytes conform transmit violate drop

  interface Ethernet1/1-3
  ```

```
switchport mode trunk
switchport access vlan 1500
service-policy type qos input APIC_Policy
service-policy type qos output APIC_Rate_Limit
```